

Antagonistic Structural Patterns in Complex Networks

María J. Palazzi,¹ Javier Borge-Holthoefer,¹ Claudio J. Tessone,² and Albert Solé-Ribalta¹

¹*Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Catalonia, Spain*

²*URPP Social Networks, Universität Zürich, Switzerland*

Identifying and explaining the structure of complex networks at different scales has become an important problem across disciplines. At the mesoscale, modular architecture has attracted most of the attention. At the macroscale, other arrangements –e.g. nestedness or core-periphery– have been studied in parallel, but to a much lesser extent. However, empirical evidence increasingly suggests that characterizing a network with a unique pattern typology may be too simplistic, since a system can integrate properties from distinct organizations at different scales. Here, we explore the relationship between some of those organizational patterns: two at the mesoscale (modularity and in-block nestedness); and one at the macroscale (nestedness). We analytically show that nestedness can be used to provide approximate bounds for modularity, with exact results in an idealized scenario. Specifically, we show that nestedness and modularity are antagonistic. Furthermore, we evince that in-block nestedness provides a parsimonious transition between nested and modular networks, taking properties of both. Far from a mere theoretical exercise, understanding the boundaries that discriminate each architecture is fundamental, to the extent modularity and nestedness are known to place heavy constraints on the stability of several dynamical processes, specially in ecology.

PACS numbers: 89.65.-s, 89.75.Fb,

The detection and identification of emergent structural patterns has been a main focus in the development of modern network theory. Such interest is not surprising, because these arrangements lie at the core of the discipline as one of the keys to the origins –which are the assembly rules that led to an observed pattern?– and dynamics –how is the system’s activity constrained by the structure?– of a network. In addition to these essential questions, the identification of structural signatures is a difficult task *per se*, which explains as well why so much attention has been put on the technical problem.

Undoubtedly, in this context, modularity [1, 2] stands out: the organization of a network as a set of cohesive subgroups has, by far, concentrated most of the efforts [3–7]. Modular architecture is widespread [8–12] and responds to the intuition that similar elements in a complex system tend to flock together. However, there are other architectural principles beyond community structure which may play more important roles. In ecology, for instance, scholars have faced the need to define and quantify other patterns, demonstrated to be more relevant in certain scenarios. This is the case of nestedness [13, 14], a concept that has been crucial to understand the stability and diversity of ecological systems [15, 16]. Other, more intricate, possibilities have also been explored [17, 18], like core-periphery structures [19, 20] –and its extension to the mesoscale [21]– are good examples of comparatively less studied architectures.

In these settings, the accent has been mainly placed on designing heuristics and improving algorithms [12, 20, 22]; understanding the dynamical constraints that those patterns impose [23, 24]; or describing plausible microscopic rules that make those patterns emerge [15, 25–27]. However, we have very limited knowledge on how different structural signatures may be intertwined, or how –if ever– they affect and limit each other. Indeed, we have examples in which two or more structural features (say, nestedness and modularity) have been jointly considered [28–31]. But such consideration overlooked to what extent the inherent constraints of one pattern

limit –or boost– the presence of the other. Furthermore, there is extensive evidence that modular and nested architectures play a critical role in relation to the stability of the dynamics of ecological systems [15, 16, 32, 33], economics [34, 35] and social sciences [31]. Thus, understanding the possibility of coexistence of these structural patterns may shed light on the dynamical trade-offs that either arrangement can facilitate.

In [31], the authors observe that modularity and nestedness exhibit an anti-correlated behavior, suggesting that, at least in empirical data, these arrangements hardly coexist. In this work, we depart from this shallow evidence to first explore experimentally, and then analytically, the structural relationship between three structural patterns: nestedness on the macroscopic side, modularity and in-block nestedness at the mesoscale. We analytically characterize these measures in an idealized family of networks, which allows us to precisely derive to what extent the macroscale organization places strict bounds to the emergence of mesoscale patterns. Eventually, we provide soft bound estimations for less restricted scenarios, *i.e.* real networks.

In a perfectly nested network, the set of neighbors of lower degree nodes are a subset of those with larger degree [36]. Such network is typically represented by a presence-absence matrix a_{ij} , and the degree of nestedness \mathcal{N} , can be formally defined as [37]

$$\mathcal{N} = \frac{2}{N_T} \left\{ \sum_{ij}^{N_T} \left[\frac{O_{ij} - \langle O_{ij} \rangle}{k_j(N_T - 1)} \Theta(k_i - k_j) \right] \right\}, \quad (1)$$

where $O_{ij} = \sum_k a_{ik}a_{jk}$ quantifies the overlap between nodes i and j ; k_i corresponds to the degree of node i ; and $\Theta(\cdot)$ is the Heaviside step function, that ensures that O_{ij} has a positive contribution when $k_i \geq k_j$. Additionally, O_{ij} is conveniently corrected by a null model that discounts the expected overlap if links were drawn randomly, $\langle O_{ij} \rangle$. N_T is the size of the network.

A modular structure is a rather ubiquitous mesoscale structural organization in which nodes are organized forming groups, i.e. devoting many links to nodes in the same group, and fewer links towards nodes outside. One of the most popular methods to identify communities is through the maximization of the modularity Q [1]. The original equation can be rewritten as

$$Q = \sum_{c=1}^B \left[\frac{l_c}{L} - \left(\frac{d_c}{2L} \right)^2 \right], \quad (2)$$

where B is the number of communities, L is the total number of links in the network, l_c is the total number of links in community c , and d_c is the sum of the degrees of all nodes in such community.

The possibility of a combined nested-modular organization has been debated in different contexts [37–39]. One conceivable form of coexistence is in terms of hybrid structures, as described by Lewinsohn *et al.* in [17]. The general layout of these networks is modular, but interactions within each module (or block) are expected to be nested. In contrast, modularity makes no assumption on the internal organization of communities. Worth highlighting, this hybrid structure reframes nestedness, originally a macroscale feature, to the mesoscopic level – it can now be interpreted as an in-block nested structure with $B = 1$. The degree of in-block nestedness of a network \mathcal{I} [37] can be computed as

$$\mathcal{I} = \frac{2}{N_T} \left\{ \sum_{i,j}^{N_T} \left[\frac{O_{ij} - \langle O_{ij} \rangle}{k_j(C_i - 1)} \Theta(k_i - k_j) \delta(\alpha_i, \alpha_j) \right] \right\}. \quad (3)$$

Here, α_i represents the community node i belongs to, and C_i its size. $\delta(\alpha_i, \alpha_j)$ corresponds to the Kronecker delta, equal to one only if nodes i and j belong to the same community.

In order to experimentally assess the dependence between the different presented structural measures, we rely on a probabilistic network generation model [37]. This model is able to generate structures (with and without noise) that smoothly interpolate between the different structural patterns of interest by recourse of 4 parameters: the number of modules B , the fraction of inter-modular links μ (inter-block noise), the fraction of links outside the perfect nested structure p (intra-block noise), and the shape of the nested structure ξ (cf. Supplemental Material).

We have measured \mathcal{N} , and optimized Q and \mathcal{I} for more than 2×10^5 networks on a wide range of parameters: $B \in [1, 9]$; $\xi \in [1.5, 7]$; $p \in [0, 0.6]$; and $\mu \in [0, 0.6]$. Restricting p and μ to 0.6 allows us to introduce a high level of noise while still preserving some underlying mesoscale structure. For modularity and in-block nestedness optimization, we have used the extremal optimization algorithm [3], adapted to the corresponding objective functions (Eqs. 2 and 3, respectively). Community size of $N_B = 50$ was assumed, so as we add communities we are also increasing the size of the network, the total number of nodes being $N_T = BN_B$. Fixing N_T and reducing the size of the communities while increasing B produces equivalent results.

Figure 1 shows the results represented over four ternary heatmap plots. A ternary plot, or simplex, is a three-variable diagram in which the sum of the variables is a constant –1 in this case– (for further details cf. Supplemental Material). Panel (a) shows a density plot of the generated networks over the simplex, and the colorbar indicates the amount of networks in each bin of the ternary plot. Relying on these results, it is apparent that the most frequent architecture is predominantly modular. This is expected since most generated networks have blocks ($B > 1$), and in-block nested networks are more restrictive, in terms of internal organization, than modular ones. The color code in panels (b), (c) and (d) reports the average absolute value of \mathcal{N} , Q and \mathcal{I} . Dashed black lines have been added as visual aids to evince *dominance* regions. A quick glance already shows that the highest values of \mathcal{N} and Q never overlap, while \mathcal{I} bridges between them. This is a valuable insight for the analytical results in the remainder of the article.

Another outstanding feature in Fig. 1 is the existence of sharp boundaries in the ternary plots. The first boundary, F_1 in panel (a), is induced by the definitions of \mathcal{N} and \mathcal{I} : as stated, Eq. 3 reduces to Eq. 1 when $B = 1$. Translated to coordinates on the simplex, F_1 simply reflects that the contribution of \mathcal{N} is always equal or smaller than the contribution of \mathcal{I} , $f_{\mathcal{N}} \leq f_{\mathcal{I}}$. More interesting, however, is the existence of F_2 , which suggests that there is an inherent limit that prohibits in-block nestedness to dominate further over Q . On close inspection (see Fig. S3 of the Supplemental Materials), networks which map onto F_2 have high values of ξ and very low values of p and μ . We build on this finding to construct our analytical approach below.

The specific configuration of parameters along F_2 in Fig. 1 points at a well-defined family of network configurations: a ring of star graphs, G_s hereafter. Indeed, an extreme shape parameter ($\xi \rightarrow \infty$), perfectly nested intra-block structure ($p = 0$) and minimum inter-block connectivity ($\mu \approx 0$) to guarantee a single giant component, render a network model which depends only on B , ranging from a single star ($B = 1$) to a set of stars ($B > 1$) connected with a single link through their central nodes. In other words, G_s provides the closest compatible network architecture for the boundary F_2 . Given a ring of star graphs with B communities and N_B nodes per community, we can analytically derive the exact values \mathcal{N} , \mathcal{I} and Q .

Nestedness. We obtain the analytical expression for \mathcal{N} from the expression in Eq. 1. The pair overlap of a generalist node (the center of each star subgraph), g , with a specialist node (periphery of a star), s , is $O_{gs}/k_s = 1$ if g and s belong to the same star (and 0 otherwise). For all those pairs (regardless of the star they belong to), the null model contribution is $\langle O_{gs}/k_s \rangle = (N_B + 1)/BN_B$. We can obtain in a similar way the terms for the generalist-generalist pairs between stars. Summing up all the contributions, the final expression for \mathcal{N} is:

$$\mathcal{N} = \frac{BN_B^3 - BN_B^2 - 3BN_B + B + 2N_B + 2}{BN_B(BN_B^2 + BN_B - N_B - 1)}. \quad (4)$$

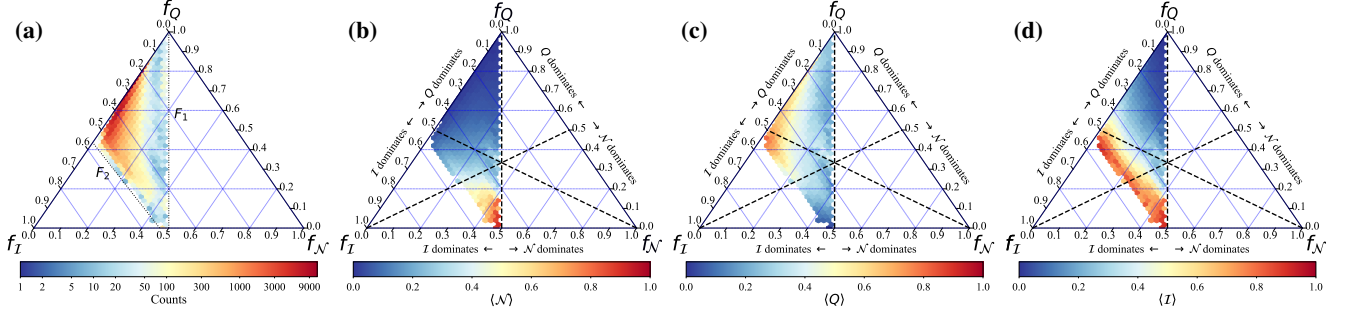


FIG. 1. Panel (a) shows the distribution of the generated networks over the ternary plot. The color bar indicates the amounts of networks in each bin. Panels (b), (c) and (d) show the average absolute value of \mathcal{N} , Q and \mathcal{I} , respectively.

Modularity. While the optimal partition for an arbitrary network cannot be easily obtained, this is not the case for G_s where each star in the ring forms a community. Thus, we can easily derive the contribution of each star to the total Q following Eq. 2. The first element is $l_c = N_B - 1$. The second element (the amount of links of the network) includes links within and between communities, $L = B(N_B - 1) + B = BN_B$. The last term, the sum of the degrees of all the nodes in community c , corresponds to $d_c = 2N_B$. Assembling these, we obtain the modularity of G_s as

$$Q = B \left[\frac{N_B - 1}{BN_B} - \left(\frac{2N_B}{2BN_B} \right)^2 \right] = 1 - \frac{1}{N_B} - \frac{1}{B}, \quad (5)$$

which is equivalent to the general expression derived in [40].

In-block nestedness. The derivation of \mathcal{I} resembles that of \mathcal{N} , with the difference that only nodes within the same community contribute; thus, all stars have the same contribution. Focusing now on each star, we have only two contributing terms to the sum: the pair overlap between specialist nodes, s , and the pair overlap of the generalist node, g , with the specialists. In both cases, the contribution is 1. The null model corrections are $\langle O_{gs} \rangle = k_g k_s / BN_B = (N_B + 1) / BN_B$ and $\langle O_{ss} \rangle = k_s k_s / BN_B = 1 / BN_B$. Finally, the size of the communities is $C_g = C_s = N_B$. Replacing all the contributions in Eq. 3, we obtain

$$\mathcal{I} = 1 - \frac{3}{BN_B} - \frac{2}{N_B}. \quad (6)$$

All the expressions presented above were obtained considering a closed ring, on which the number of intercommunity links is B . For the cases $B = 1$ and $B = 2$, the number of intercommunity links is $B - 1$ and the degree of the generalist nodes is $k_g = N_B - 1$ and $k_g = N_B$, respectively (cf. Supplemental Materials for details on these cases).

We now focus on the bounds that \mathcal{N} and Q impose on each other in some important limits. These correspond to scenarios in which the number of blocks, B , and the size of the blocks, N_B , tend to ∞ .

We start with $N_B \rightarrow \infty$. In this case, Eqs. 4 and 5 reduce to

$$\lim_{N_B \rightarrow \infty} \mathcal{N} = \frac{1}{B}, \quad \lim_{N_B \rightarrow \infty} Q = 1 - \frac{1}{B}, \quad (7)$$

which implies that, under these circumstances, \mathcal{N} and Q are complementary—in accordance with the empirical results in [31]. This result proves analytically the antagonism that exists between these two structural patterns.

With respect to the case $B \rightarrow \infty$, Eqs. 4 and 5 turn now

$$\lim_{B \rightarrow \infty} \mathcal{N} = 0, \quad \lim_{B \rightarrow \infty} Q = \frac{N_B - 1}{N_B}. \quad (8)$$

These results fit the expectation that, with increasing B , the negative contribution of non-overlapping nodes through the null model overcomes the decreasing positive contributions realized.

Finally, with respect to in-block nestedness, the analytical calculations in both limits yield

$$\lim_{N_B \rightarrow \infty} \mathcal{I} = 1 \quad \lim_{B \rightarrow \infty} \mathcal{I} = \frac{N_B - 2}{N_B} \quad (9)$$

Noteworthy, the generated synthetic networks follow closely the predictions in limiting cases: Fig. 2 reports the analytical estimation of \mathcal{N} , Q and \mathcal{I} (Eqs. 4-6, symbols in the figure) against B , and the numerical results for networks generated under different parameters. As the generated networks deviate from the ring of stars G_s (i.e. $p > 0$ and $\mu > 0$), results show a worse fit to the analytical prediction. The mutual bounds that \mathcal{N} and Q impose on each other are obvious, observing a perfectly anti-correlated behavior between nestedness and modularity. Finally, as the networks transition from a nested ($B = 1$) to a modular ($B > 1$) architecture, the values of in-block nestedness remain very high and almost constant.

The previous results open a new front to understand the co-occurrence of macro- and mesoscale patterns in complex networks. Complementary to the inherent limits of Q [40, 41], we have now evidence that a certain connectivity arrangement (i.e. nestedness) places hard limits to modularity, at least in extreme settings. This certainty paves the way to obtain estimations for Q prior to computationally costly endeavors: indeed, relaxing those conditions to realistic parameters, soft bounds for Q may be defined. The derivation of these bounds are presented below.

We start from a G_s of size $N_T = BN_B$. From here, Eq. 4 can be rewritten in terms of N_T and B , and an estimation on the number of blocks as a function of \mathcal{N} and N_T can be obtained, i.e. $B^* = f(\mathcal{N}, N_T)$. The upper bounds for Q and

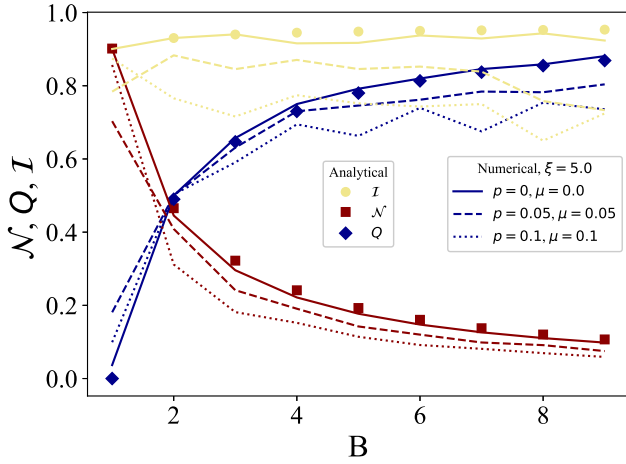


FIG. 2. Comparison of the analytical (symbols) and numerical (lines) values of \mathcal{N} , Q , \mathcal{I} with respect to B . All the calculations were performed by taking $N_B = 50$ and $\xi = 5$. The values for p and μ parameters are indicated in the plot legend.

\mathcal{I} are thus readily available, applying B^* to Eqs. 4-5, i.e. assuming that the network structure lies on the boundary F_2 . With the actual measure of \mathcal{N} , and upper estimations for Q and \mathcal{I} , we can obtain the relative fraction that each measure contributes to the ternary plot along the F_2 boundary, $f_{\mathcal{N}}^{\uparrow}$, f_Q^{\uparrow} and $f_{\mathcal{I}}^{\uparrow}$.

To obtain the lower bounds for Q , we observe from Fig. 1(b) that, if we move from boundary F_2 to boundary F_1 on the f_Q -axis direction, i.e. horizontally in Fig. 1, the \mathcal{N} values are approximately constant with respect to the contributions of Q (f_Q^{\uparrow} and f_Q^{\downarrow}). This allows us to make an approximation for the contributions of Q in the ternary plot at F_1 as $f_Q^{\downarrow} \approx f_Q^{\uparrow}$. Additionally, we know that $\mathcal{N} = \mathcal{I}$ at boundary F_1 . Thus,

$$f_Q^{\downarrow} = \frac{Q}{Q + \mathcal{I} + \mathcal{N}} = \frac{Q}{Q + 2\mathcal{N}}, \quad (10)$$

from which a lower bound for Q can be obtained.

Figure 3 shows the values of Q as a function of \mathcal{N} for the previous synthetic ensemble ($\sim 2 \times 10^5$ networks; panel (a), grey dots); and 57 real unipartite networks (panel (b), red dots) [37]. In panel (a), the values of the theoretical upper and lower bounds are plotted in colors, the color bar indicating the network size. Our approximation of Q bounds is in good agreement with actual values obtained after optimization: most of

the optimized Q values lie within the estimated soft bounds. Despite the wide range of parameters—clearly far from limiting cases—estimated upper bounds behave like $Q = 1 - \mathcal{N}$ almost perfectly. While these bounds are trivial for $\mathcal{N} \approx 0$, we observe that intermediate values of nestedness provide relevant information about the possible mesoscale organization of the network. Q values above the upper bound correspond to networks with a single community $B = 1$ and perfectly nested structure, $p = 0$ (see Fig. S4). These networks—less than 0.1% of the total—are dense enough to allow a partition with $B > 1$ where the nodes of higher degree are gathered in a block, resulting in values of Q larger than expected [37]. Values below the lower bound approximation are more numerous—although still a small fraction of the total. This imprecision shows that there is room to improve the underlying assumption, i.e. that \mathcal{N} values are constant with respect to the contributions of Q . In the same spirit, upper and lower bounds for \mathcal{I} can be as well approximated from the actual value of \mathcal{N} (see Fig. S5). For the sake of completeness, Q - \mathcal{I} scatter plots are shown in Fig. S6, where we see that \mathcal{I} and Q can coexist, i.e. there is no clear map from one to the other. Remarkably, bound estimation for real networks in Fig. 3(b) closely follows the approximation for synthetic networks: the inferior and superior trends of black dots are the predicted lower and upper bounds. It is worth highlighting, that bound estimation—which has a very low computational cost—renders non-trivial information for some networks ($\mathcal{N} \gtrsim 0.2$).

While the study of macro- and mesoscale arrangements in complex networks has been studied in depth, we know little about how they affect each other. Understanding and, above all, quantifying such pattern interactions becomes necessary for many reasons. First, because empirical evidence suggests the concurrence of more than one pattern within the same network [28–31]. Second, because a preliminary approximation of the mesoscale structural features of a network is appealing, at the face of prohibitive costs to analyze very large amounts of data. Further, the interplay between nestedness and modularity is thought fundamental to decipher the dynamical behavior of many empirical systems (like ecological, economic, and technological networks among others). In this work, we have quantified, numerically and analytically, the interference between nestedness (at the macroscale) and modularity and in-block nestedness (at the mesoscale). We show that modularity and nestedness are antagonistic architectures: the growth of one implies the decline of the other, and bounds to modularity can be estimated even in far from idealized settings. Intermediate nested-modular regimes are possible, pointing directly at in-block nested structures as the natural transition between the other two.

- [1] M. E. Newman and M. Girvan, Physical review E **69**, 026113 (2004).
- [2] M. E. Newman, Proceedings of the national academy of sciences **103**, 8577 (2006).
- [3] J. Duch and A. Arenas, Physical review E **72**, 027104 (2005).
- [4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre,

Journal of Statistical Mechanics: theory and experiment **2008**, 10008 (2008).

- [5] M. E. Newman, Physical review E **70**, 056131 (2004).
- [6] E. A. Leicht and M. E. Newman, Physical review letters **100**, 118703 (2008).
- [7] M. J. Barber, Physical Review E **76**, 066102 (2007).

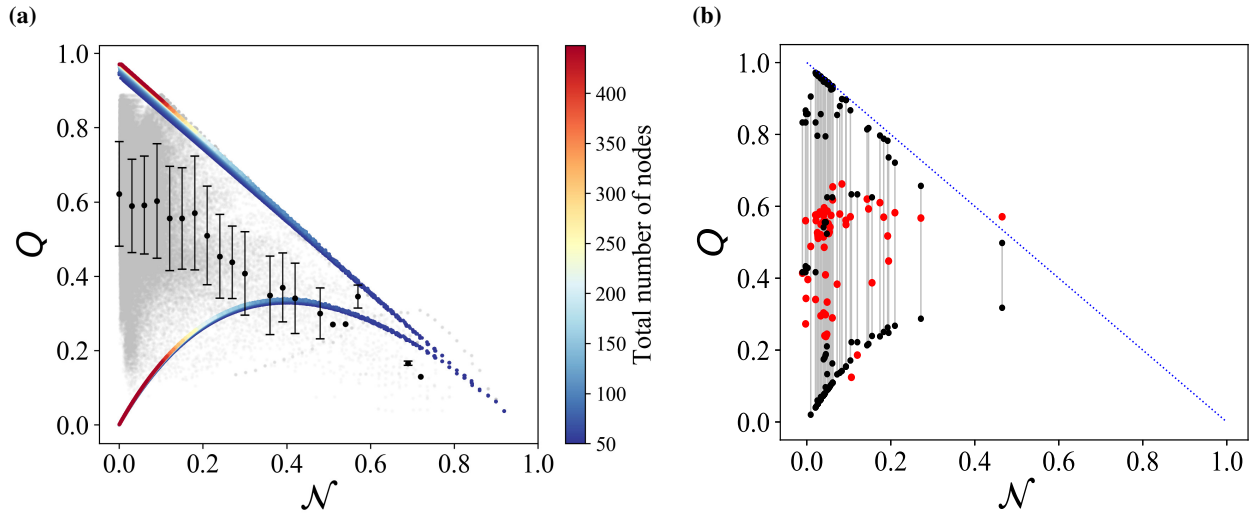


FIG. 3. Panel (a) show the values of Q obtained after optimization (grey dots), plotted against N for over 2×10^5 generated networks and panel (b) for the set of unipartite social network (57 networks) analyzed in [37]. Colored upper and lower bounds of Q have been obtained from Eqs. 5-6 and $B^* = f(N, N_T)$. The color bar indicates the network's size.

- [8] W. W. Zachary, *Journal of Anthropological Research* **33**, 452 (1977).
- [9] R. Guimerà and L. A. N. Amaral, *Nature* **433**, 895 (2005).
- [10] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen, *Physical Review Letters* **90**, 148701 (2003).
- [11] L. A. Adamic and N. Glance, in *Proceedings of the 3rd international workshop on Link discovery* (ACM, 2005) pp. 36–43.
- [12] S. Fortunato, *Physics Reports* **486**, 75 (2010).
- [13] W. Atmar and B. D. Patterson, *Oecologia* **96**, 373 (1993).
- [14] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen, *Proceedings of the National Academy of Sciences* **100**, 9383 (2003).
- [15] E. Thébaud and C. Fontaine, *Science* **329**, 853 (2010).
- [16] U. Bastolla, M. A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, and J. Bascompte, *Nature* **458**, 1018 (2009).
- [17] T. M. Lewinsohn, P. Inácio Prado, P. Jordano, J. Bascompte, and J. M. Olesen, *Oikos* **113**, 174 (2006).
- [18] M. Almeida-Neto, P. R. Guimarães Jr., and T. M. Lewinsohn, *Oikos* **116**, 716 (2007).
- [19] S. P. Borgatti and M. G. Everett, *Social Networks* **21**, 375 (2000).
- [20] P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, *SIAM Review* **59**, 619 (2017).
- [21] S. Kojaku and N. Masuda, *Physical Review E* **96**, 052313 (2017).
- [22] J.-H. Lin, C. Tessone, and M. Mariani, *Entropy* **20**, 768 (2018).
- [23] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente, *Physical Review Letters* **96**, 114102 (2006).
- [24] S. Allesina and S. Tang, *Nature* **483**, 205 (2012).
- [25] C. Y. J. Leung and J. S. Weitz, *Physical Review E* **93**, 032303 (2016).
- [26] M. D. König, C. J. Tessone, and Y. Zenou, *Theoretical Economics* **9**, 695 (2014).
- [27] T. Verma, F. Russmann, N. Araújo, J. Nagler, and H. J. Herrmann, *Nature Communications* **7**, 10441 (2016).
- [28] J. M. Olesen, J. Bascompte, Y. L. Dupont, and P. Jordano, *Proceedings of the National Academy of Sciences* **104**, 19891 (2007).
- [29] M. A. Fortuna, D. B. Stouffer, J. M. Olesen, P. Jordano, D. Mouillot, B. R. Krasnov, R. Poulin, and J. Bascompte, *Journal of Animal Ecology* **79**, 811 (2010).
- [30] C. O. Flores, J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz, *Proceedings of the National Academy of Sciences* **108**, E288 (2011).
- [31] J. Borge-Holthoefer, R. A. Baños, C. Gracia-Lázaro, and Y. Moreno, *Scientific Reports* **7**, 41673 (2017).
- [32] D. B. Stouffer and J. Bascompte, *Proceedings of the National Academy of Sciences* **108**, 3648 (2011).
- [33] S. Allesina, J. Grilli, G. Barabás, S. Tang, J. Aljadeff, and A. Maritan, *Nature Communications* **6**, 7842 (2015).
- [34] S. Bustos, C. Gomez, R. Hausmann, and C. A. Hidalgo, *PloS one* **7**, e49393 (2012).
- [35] S. Saavedra, D. B. Stouffer, B. Uzzi, and J. Bascompte, *Nature* **478**, 233 (2011).
- [36] Without loss of generality, in the following we consider purely unipartite networks. The extension to bipartite ones is trivial.
- [37] A. Solé-Ribalta, C. J. Tessone, M. S. Mariani, and J. Borge-Holthoefer, *Physical Review E* **96**, 062302 (2018).
- [38] C. O. Flores, S. Valverde, and J. S. Weitz, *The ISME Journal* **7**, 520 (2013).
- [39] S. J. Beckett and H. T. Williams, *Interface Focus* **3**, 20130033 (2013).
- [40] S. Fortunato and M. Barthélemy, *Proceedings of the National Academy of Sciences* **104**, 36 (2007).
- [41] A. Lancichinetti and S. Fortunato, *Physical Review E* **84**, 066122 (2011).

Supplemental Material for “Antagonistic Structural Patterns in Complex Networks”

María J. Palazzi¹, Javier Borge-Holthoefer¹, Claudio J. Tessone² and Albert Solé-Ribalta¹
¹*Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Catalonia, Spain*
²*URPP Social Networks, Universität Zürich, Switzerland.*
 (Dated: November 9, 2018)

The following section provides a detailed description of the network generation model employed to carry out the experimentally explore the structural relationship between nestedness, modularity and in-block nestedness. The model is a version of the one developed by Solé-Ribalta *et al.* in [1] modified to generate size increasing networks with a fixed block size, instead of networks with fixed size, as in the original formulation.

I. SUPPLEMENTAL SECTION I: PROBABILISTIC MODEL FOR THE SYNTHETIC NETWORK GENERATION

We generated synthetic in-block nested networks employing the benchmark graph model introduced in [1]. The model is implemented in terms of links probabilities, and asks for four parameters: the number of modules B , the fraction of inter modules links μ , the fraction of links outside the perfect nested structure p , and the shape parameter ξ , that represents the slimness of the nested structure.

The perfect nested structure is generated using a function inspired in the p -norm ball equation, written as

$$f_n(x) = 1 - (1 - x^{1/\xi})^\xi, \quad (1)$$

where $\xi \in [1, \infty)$ and $x \in [0, 1]$. The perfect nested structure is constructed by adding a link into each matrix position whose center lies above the curve in Eq. 1, resembling an upper left triangle. Afterwards, starting with a fixed community size and given a number of communities B , we construct the adjacency matrix of a network by building $\lfloor B \rfloor$ blocks of size n_r and a remaining block of size $\{B\}n_r$, such that, the total size of the network is $N_r = Bn_r$, forming a block diagonal matrix^{1,2}.

The probability of having a link between two nodes i and j inside a block is given by

$$P(A_{ij}^c) = [(1 - p + pp_r)\Theta(jN_r - f_n(iN_r)) + p_r(1 - \Theta(jN_r - f_n(iN_r)))](1 - p_i), \quad (2)$$

The term within square brackets is related to the intra-block noise, being p the probability of removing a link from the perfect nested structure. Hence, the term $(1 - p)$ corresponds to the probability of not altering the link. The second, pp_r , corresponds to the probability of recovering a link after removal and $p_r = pE(N_r - E + pE)^{-1}$, corresponds to the probability of selecting link A_{ij} from the distribution of removed links, where E is the number of links in the network. The first two terms are restricted by the Heaviside function Θ , to the region of perfect in-block nestedness. Finally, the term $(1 - p_i)$ corresponds to the probability of not removing the link in the process of generating inter-block noise, where $p_i = \mu(B - 1)/B$ and $\mu \in [0, 1]$.

Finally, the probability of a inter-block link is given by,

$$P(A_{ij}^o) = \frac{2Ep_i}{2(B - 1)N_r^2} = \frac{\mu E}{N_r^2 B}, \quad (3)$$

the numerator corresponds to the number of removed links from the blocks, and the denominator corresponds to the possible places where each of those links can be relocated.

Figure S1 shows several examples of the networks the model is able to generate. Perfectly nested networks are generated with $B = 1$ and varying values of ξ , then fixing $p = \mu = 0$. With the same settings and $B > 1$, perfect in-block nested networks are obtained. Ideal modular networks are generated fixing $\mu = 0$ and varying values of p and ξ . Parameters $p = \mu = 1$ generate Erdős-Rényi networks, regardless of B (bottom-right).

¹ $\lfloor \cdot \rfloor$ refers to the integer part function

² $\{\cdot\}$ refers to the fractional part function

and \mathcal{I} for the cases of a single star ($B = 1$) and two stars connected through their central nodes ($B = 2$). As stated in the main text, for these two situations, we have to take into account the change in the number of inter-community links and the degree of the generalist nodes.

A. Nestedness

1. A single star graph, $B = 1$

The computation of the pair overlap for the evaluation of nestedness when $B = 1$ requires only the following terms: the pair overlap of a generalist node (the center of each star subgraph), g , with the specialist nodes s which is $O_{gs}/k_s = 0$; and the pair overlap between all the specialists nodes $O_{ss}/k_s = 1$, the degree of the generalist node is $k_g = N_B - 1$ and the null model corrections $\langle O_{gs} \rangle = k_g k_s / BN_B = (N_B - 1) / BN_B$ and $\langle O_{ss} \rangle = k_s k_s / BN_B = 1 / BN_B$

$$\begin{aligned} \mathcal{N} &= \frac{2}{N_B(BN_B - 1)} \left\{ \left[-\frac{N_B - 1}{BN_B} (N_B - 1) \right] + \left[\left(1 - \frac{1}{BN_B} \right) \frac{(N_B - 2)(N_B - 1)}{2} \right] \right\} \\ &= \frac{(N_B - 2)(N_B - 1)}{N_B^2} - \frac{2(N_B - 1)^2}{N_B^2(N_B - 1)}, \end{aligned} \quad (4)$$

2. Two star graph, $B = 2$

For this scenario we have to take into account the change on the degree of the generalist node $k_g = N_B$ and additional terms such as: the pair overlap between the two generalists O_{gg}/k_g , the pair overlap between a generalist with the specialist from the other community $O_{gs_{out}}/k_s$, and the pair overlap between a specialist with the specialists from the other community $O_{ss_{out}}/k_s$. Finally, we obtain

$$\mathcal{N} = \frac{BN_B^3 - BN_B^2 - 6N_B^2 + 8N_B - 3}{BN_B^2(BN_B - 1)} \quad (5)$$

B. Modularity

1. A single star graph, $B = 1$

Starting with the equation for modularity expressed as sum over the communities (Eq. 2 of the main text), we obtain the total number of links in the network and the number of links per community for a single star graph $N_B - 1$, and the sum of the degrees of the nodes in the community $d_c = 2(N_B - 1)$. Now, we have the maximum modularity for $B = 1$.

$$Q = B \left[\frac{(N_B - 1)}{(N_B - 1)} - \left(\frac{2(N_B - 1)}{2(N_B - 1)} \right)^2 \right] = 0, \quad (6)$$

2. Two star graph, $B = 2$

When $B = 2$ the total number of links changes to $L = B(N_B - 1) + 1$ and the sum of the degrees of the nodes in the community is $d_c = 2(N_B - 1) + 1$. So we obtain

$$\begin{aligned}
Q &= B \left[\frac{(N_B - 1)}{B(N_B - 1) + 1} - \left(\frac{2N_B - 1}{2(B(N_B - 1) + 1)} \right)^2 \right] \\
&= 2 \left[\frac{(N_B - 1)}{2(N_B - 1) + 1} - \left(\frac{2N_B - 1}{2(2(N_B - 1) + 1)} \right)^2 \right] \\
&= \left[\frac{(2N_B - 2)}{2N_B - 1} - \frac{1}{2} \right],
\end{aligned} \tag{7}$$

C. In-block nestedness

1. A single star graph, $B = 1$

Once again, we know that we will have only two contributing terms to our sum; the pair overlap between specialists (s) nodes and the pair overlap of the generalist (g) node with the specialists. Additionally, we know that for this case the degree of the generalist node is $k_G = N_B - 1$ and the rest of the terms are: the number of specialists nodes $N_s = (N_B - 1)$, the null model corrections $\langle O_{g,s} \rangle = k_g k_s / BN_B = (N_B - 1) / BN_B$ and $\langle O_{s,s} \rangle = k_s k_s / BN_B = 1 / BN_B$, and the size of the communities is $C = N_B$. Finally the analytical expression for in-block nestedness when $B = 1$ reads,

$$\begin{aligned}
\mathcal{I} &= \frac{2}{N_B} \left\{ \left[\frac{-(N_B - 1) / BN_B}{(N_B - 1)} (N_B - 1) \right] + \left[\frac{1 - 1 / BN_B}{(N_B - 1)} \frac{(N_B - 2)(N_B - 1)}{2} \right] \right\} \\
&= \frac{2}{N_B} \left\{ \left[-\frac{(N_B - 1)}{N_B} \right] + \left[\frac{(N_B - 1)(N_B - 2)}{2N_B} \right] \right\},
\end{aligned} \tag{8}$$

2. Two star graph, $B = 2$

Now, for $B = 2$ we have that the degree of the generalist node is $k_G = N_B$. Substituting this term we obtain a new expression for the in-block nestedness as

$$\mathcal{I} = \frac{2}{N_B} \left\{ \left[-\frac{1}{N_B} \right] + \left[\frac{(2N_B - 1)(N_B - 2)}{4N_B} \right] \right\}, \tag{9}$$

IV. SUPPLEMENTAL FIGURE 3

Figure S3 shows the results with respect to the parameters of the probabilistic network generation model employed to perform the numerical exploration as explained in the main text (Fig 1). The model is described in detail in the Supplemental Section I. Panel (a)-(d) show the results with respect to varying values of the number of blocks (B), shape parameter (ξ), intra-block (p) and inter-block noise (μ), respectively. The color bar indicates the mean value of the respective parameter in each bin of the simplex.

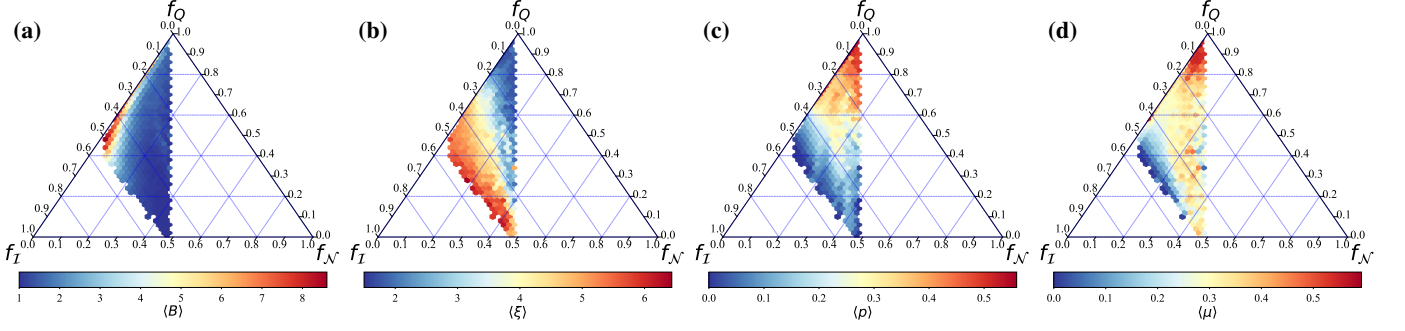


FIG. S3: Ternary plots representing results for $\sim 2 \times 10^5$ networks as in the main text. In this case, color in each bin of the simplex indicates the average number of blocks B (A); average shape parameter ξ (B); average intra-block noise p (C); and finally average inter-block noise μ (D).

V. SUPPLEMENTAL FIGURE 4

Figure S4 shows the values of Q plotted against \mathcal{N} , for the all the generated networks employed in the numerical exploration in the main text ($\sim 2 \times 10^5$). The corresponding upper and lower bounds were plotted on top. The color bar, in each case, indicates the values of the respective parameters of the probabilistic network generation model (number of blocks (B), shape parameter (ξ), intra-block (p) and inter-block noise (μ), respectively).

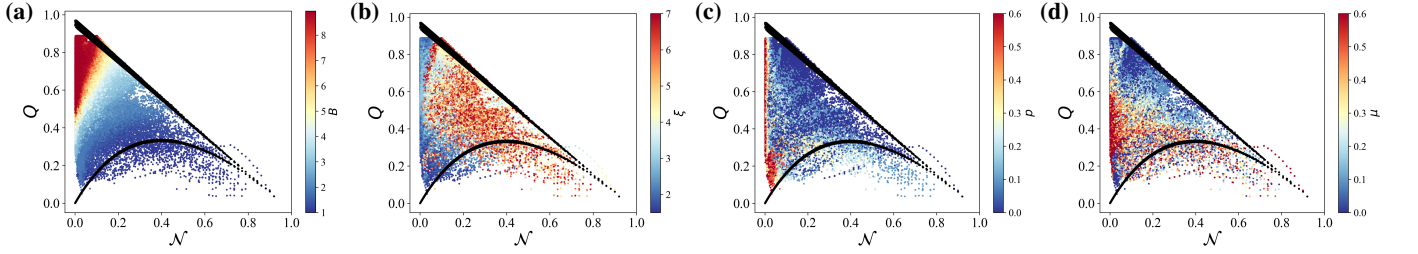


FIG. S4: Optimized values of Q plotted against \mathcal{N} , for the generated networks. The values of the corresponding upper and lower bounds were plotted on top (black dots). The color bar indicates the value of the respective parameters of the probabilistic network generation model (number of blocks (B), shape parameter (ξ), intra-block (p) and inter-block noise (μ), respectively).

VI. SUPPLEMENTAL FIGURES 5 AND 6

For the sake of completeness, we have plotted the values of \mathcal{I} against \mathcal{N} (Fig. S5) and Q against \mathcal{I} (Fig. S6).

Similar to Fig. 3A in the main text, upper and lower bounds for \mathcal{I} have been calculated in Fig. S5, taking actual measurements of \mathcal{N} as a starting point. Remarkably, none of the optimized values of \mathcal{I} violates such bounds. This is no surprise with respect to lower bounds, since \mathcal{I} reduces to \mathcal{N} when $B = 1$, thus the lower bound simply represents the hard limit $\mathcal{I} = \mathcal{N}$. But even the upper bounds, which represent an estimation, are in excellent agreement with respect to the optimized values of \mathcal{I} .

On the other hand, the main lesson from Fig. S6 is the fact that, unlike Q and \mathcal{N} , other patterns can coexist, i.e. there is no clear map between Q and \mathcal{I} .

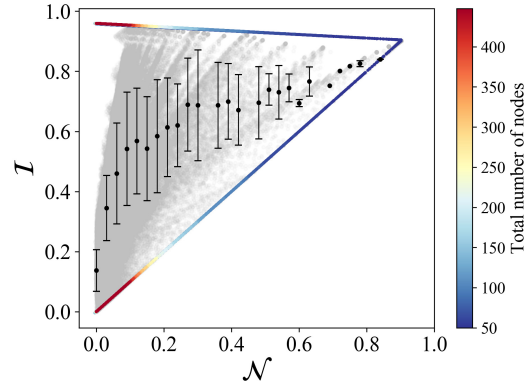


FIG. S5: Values of \mathcal{I} obtained after optimization (grey dots), plotted against \mathcal{N} for the generated networks. Upper and lower bounds of \mathcal{I} are plotted in colors. The color bar indicates the network's size

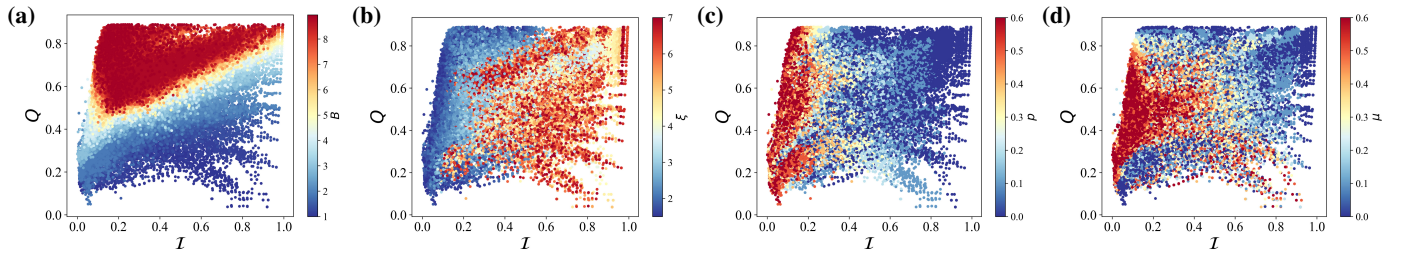


FIG. S6: Optimized values of Q plotted against the optimized values of \mathcal{I} , for the generated networks. The color bar indicates the value of the respective parameters of the probabilistic network generation model. Panel (a) show the results with respect to the number of blocks. Panel (b) correspond to the shape parameter ξ . Panels (c) and (d) correspond to the noise parameters p and μ , respectively.

-
- [1] A. Solé-Ribalta, C. J. Tessone, M. S. Mariani, and J. Borge-Holthoefer. Revealing in-block nestedness: detection and benchmarking. *Phys. Rev. E*, 97(6), 062302, 2018.